Who do you refer to? How young students with mild intellectual disability

confront anaphoric ambiguities in texts and sentences

Gema Tavares[1], Inmaculada Fajardo, Vicenta Ávila, Ladislao Salmerón, Antonio Ferrer

ERI Lectura, University of Valencia

[1] Gema Tavares is now at Department of Spanish, Europaschule Gymnasium

Westerstede

Correspondence concerning this article should be addressed to Gema Tavares,

Europachule Gymnasium Westerstede, Gartenstr. 16, 26655 Westerstede (Germany).

Email: tavares@gym-wst.de

Abstract

Along 2 experiments we tested the anaphoric pronoun resolution abilities of readers with intellectual disability in comparison with chronological and reading age-matched groups. In Experiment 1, the anaphor test of Elosúa, Carriedo and García-Madruga (2009) confirmed that readers with intellectual disability (ID) are slower than control readers resolving clitic anaphoric pronouns, especially when the use of morphological cues (e.g. gender) is necessary. In order to test if the poor performance could be due to low levels of metacognitive skills during reading, an inconsistency detection task combined with eye tracking was designed in Experiment 2. Participants read short texts with an anaphoric pronoun in the fifth sentence, either morphologically (gender) consistent or not with the information provided in the second sentence. The scores in the anaphor comprehension questions presented after the text confirmed that readers with ID are affected by the gender inconsistency but they are unable to explicitly report it and recover from it, as the number of re-fixations after reading the critical sentence suggests. As their answers to the explicit detection questions showed, the adults control group did not show any preference for morphosyntax or semantics in spite of being aware of the inconsistency. In sum, both groups of readers with and without ID are affected by inconsistencies, but ID readers do not have appropriate metacognitive skills to explicitly identify the source of the inconsistency and fix it.

Keywords: anaphora resolution; inconsistency detection; intellectual disabilities; metacognitive skills; text comprehension

Who do You Refer to? How Young Students with Mild Intellectual Disability

Confront Anaphoric Ambiguities in Texts and Sentences

An anaphor is a linguistic reference to an antecedent piece of text (Rademaker & Haeusler, 2008). Though the form of an anaphor is varied: repetition (e.g. *the cake* and the coffee are in the table. *The cake* is delicious.), pronouns (*the cake - it*), lexical co-reference (*the cake - the dessert*) or even ellipsis (*the cake - Ø*), its function is always to maintain text cohesion by recalling previously mentioned entities (antecedents) without the need of merely repeating them. Struggling readers such as readers with Intellectual Disability (ID), who are the target of the present study may lack the adequate skills to identify and understand an anaphor, which is core for text comprehension.

How an anaphor is resolved depends on intralinguistic features that is, related to the text itself: word length (e.g. number of characters or syllables), word frequency of the anaphor antecedent (e.g. *domicile* vs. *house* as possible antecedents of the anaphoric pronoun *it*), distance between the antecedent and the anaphor (e.g. number of sentences or words between them) and kind of anaphor (e.g. pronoun, repeated name or ellipsis) (Arnold, Brown-Schmidt, & Trueswell, 2007; Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Cacciari, Carreiras, & Barbolini, 1997; Crawley, Stevenson, & Kleinman, 1990; Frederiksen, 1981; Garvey, Caramazza, & Yates, 1975; Gelormini-Lezama & Almor, 2011; Gordon, Grosz, & Gilliom, 1993; Järvikivi, van Gompel, Hyönä, & Bertram, 2005) as well as extra linguistic factors (related to the reader), mainly working memory (the memory system in charge of temporarily storing and managing the information, Baddeley, 1992) and metacognitive skills such as planning, checking and revising strategies during reading comprehension (Cain, Oakhill, Barnes, & Bryant, 2001; Cain & Oakhill, 1999; Ehrlich, Remond, & Tardieu, 1999; Long & De Ley, 2000; Yuill & Oakhill, 1988) .

Apart from the semantics of the antecedent, in some languages such as Spanish pronominal anaphors bear a grammatical load for gender and number that helps to establish the link with the antecedent. Thus, finding a suitable antecedent for a pronoun can be accomplished through either (or both) of these grammatical features. Imagine a sentence like "Peter (m) gave one more coin to Thomas (m). He (m) had too many". Here, both *Peter* and *Thomas* could be the referred antecedent of the pronoun *he*. However, the context indicates that the most likely situation would be Peter giving out a coin as the result of having too many. In "Maria (f) gave one more coin to Thomas (m).He (m) had too many" *Maria* is the one awarding *Thomas* with a coin. A first probabilistic analysis will lead us to the same conclusion as in the former example. Nonetheless, the gender load of the pronoun indicates that the antecedent should be a masculine entity, becoming Thomas the only possible candidate. As one might expect, this duality is sometimes problematic resulting in the reader finding troubles to attach the pronoun to its proper antecedent.

There exist some antecedents on this psycholinguistic issue, as for instance the study of Oakhill and Yuill (1986) that explored the inferences drawn during a pronoun resolution task performed by skilled and low-skilled 7-8 year olds, who read two-clause subordinate sentences where the proper names were either of the same or different gender (e.g. Peter lent ten pence to Liz because she was very poor). In a first experiment, the subordinate clause was introduced by a pronoun referring to either the subject or the object of the main clause; in the second one, a gap was presented instead of the pronoun for participants to fill it in. Additionally, a comprehension question was presented right after the stimulus in the first experiment. Results showed that low-skilled readers encountered more difficulties drawing inferences about pronominal antecedents than their skilled peers, even when there was a gender clue to link the pronoun to the correct antecedent. These difficulties appeared especially when inferences were complex or included a higher memory load (i.e. the proper

names were not given again in the question, so the reader should remember them). The authors found no interaction between level of reading comprehension and memory load, as both groups of readers performed better in the simpler conditions, always with lower accuracy rates in the low-skilled readers. An explanation in terms of metacognitive skills is suggested: low-skilled readers sometimes decide not to go back to the disambiguating information trusting thus in their representation of the text, even though sometimes this is not accurate. Skilled readers, on the contrary, go back in the text when they have doubts about their own representation. In addition, it was also argued that low-skilled readers did not pay enough attention to cues such as the gender of the antecedent.

Megherbi and Ehrlich (2005) also corroborated the conflict of ambiguous pronouns and inference making in spoken language. In this case they followed the hypothesis that 7-8 year children presenting problems in pronoun resolution by reading, should also struggle with them in spoken language. In an anaphoric resolution task children had to decide whether the final word of a sentence should be a masculine or feminine pronoun (e.g. according to the fairy tale, Cinderella put on a beautiful dress to meet the handsome prince. She danced with… him / her). Their findings showed a higher influence of the verb bias (tendency to link the pronoun to the subject or the object) on low-skilled readers, meaning that they trusted the semantics of the verb more than the gender of the pronoun itself in order to disambiguate it. On the other hand, in line with Oakhill and Yuill (1986)'s findings skilled comprehenders took advantage of the gender cues of the pronoun over verb bias, allowing them to get an extra benefit when there existed no conflict between both linguistic features.

The results of these two studies agree with the findings of Elosúa, Carriedo and García-Madruga (2009). Elosúa et al. (2009) investigated the resolution of clitic pronominal anaphora when morphosyntax (e.g. gender and number clues) and semantics come to conflict. Clitics are a kind of pronouns devoted to act as a direct or indirect object, in the case of

Spanish from either a pre-verb or a post-verb position (e.g. pre-verb position: Lisa *lo* cogió

mientras *lo* miraba; Lisa took it while staring at it; post-verb position: La madre gritó: "¡Lisa,

cóge*lo*!"; the mother shouted: "Lisa, take it!"). In their experiment 1b, Elosúa et al. (2009)

compared the performance, in terms of speed and accuracy, of undergraduates in an anaphora

resolution task in which the anaphor could be either morphosyntactically or semantically

resolved. Results indicated that participants performed equally well in the two types of

disambiguation, morphosyntactic and semantic, though it took significantly longer to resolve

the semantic ambiguity than the morphosyntactic one. The authors also found a facilitative

effect of morphosyntax over semantics, very likely due to the nature of the analysis required

to benefit from them and the cognitive effort needed: using a semantic strategy implies the

performance of a deeper and more resource-demanding analysis, while following the

morphosyntactic clues only requires a surface analysis of the lexical units and the grammar

rules.

These antecedents explored anaphora resolution by high- and low-skilled readers but

not by youngsters with intellectual disabilities. Though the literature on this matter and

population is scarce, a brief revision of the antecedents is made in the next paragraphs as we

consider that assuming that readers with ID are low-skilled readers is a fair simply and

inaccurate assumption.

Our own previous research confirmed that young readers with ID experience problems

when they have to use and comprehend cohesive elements like connectives (e.g. but, besides,

for that reason) in text. More specifically, Fajardo, Tavares, Ávila, and Ferrer (2013) found

that readers with ID were less likely to select the target connective in a cloze task than

chronologically age-matched readers (mean age = 21 years). Conclusions by Fajardo et al.

(2014) added that the number of co-references (anaphors) contained in journalistic texts

predicted negatively the literal comprehension (Fajardo et al., 2014). In other words, the

higher the number of co-references (whatever the type) the lower the scores in literal

comprehension questions by poor readers with ID.

To our knowledge, the only previous existing literature regarding the processing of

pronouns was conducted with individuals with a specific aetiology related to ID, such as

Down syndrome (DS). For instance, Perovic (2006) focused on how DS individuals (17-21

years) resolved reflexive pronouns in comparison with English typically developing children

(5-11 years). Though participants from both groups were matched in receptive vocabulary,

DS participants presented problems in the resolution of reflexives (herself) but not in

pronouns (her), following the opposite pattern of acquisition to Typically Developed (TD)

children. In other words, the acquisition of such co-referentials by DS was not delayed but

qualitatively different, since previous research (Jakubowicz, 1991; Santiuste, 1997) had

argued that reflexives in TD children are acquired earlier than pronouns.

In sum, previous research has evidenced that individuals with ID could present singular

patterns of pronoun acquisition in comparison with regular readers. In addition, they might

find difficulties using cohesive elements such as connectives and co-references during

reading, though there are no antecedents in the literature testing how anaphors, and

particularly pronominal anaphors, are resolved by this population. In order to overcome this

lack of research we designed Experiment 1, where two groups of young readers, one

conformed by ID students and the other by TD undergraduates were asked to perform the

abovementioned anaphor task of Elosua et al. (2009). The methodological details and

operative hypotheses of this experiment are exposed in the next section. In order to test if the

observed differences between groups in Experiment 1 could be due to metacognitive factors

during reading, an inconsistency detection task combined with eye tracking was designed in

Experiment 2.

Experiment 1

Method

Participants. *Experimental group*. Twenty-nine students with ID (15 males and 14 females with an average age of 19.72, *SD* = 2.2, ranging from 16 to 24) recruited at *Camí Obert* (Valencian Community, Spain) and participated voluntarily in the study. Cami Obert is a Vocational Training Center for People with Special Needs (Associated to APSA, an Association for People with Intellectual Disability of Alicante, Spain). According to the PCPI (the Spanish initials for the First Professional Qualification Programme) Regulating Ordinance released in 2008 (Valencian Community), this kind of teaching programs last 2 years and only admits young people (aged from 16 to 21 years) with an official accreditation of disability who has fulfilled the period of mandatory scholarship (10 years in Spain) and with enough personal and social autonomy as to allow them to participate in the learning process as well as to access and maintain a job. The vehicular language of Camí Obert is Spanish.

The inclusion criteria for participants with ID were:

 (a) borderline to moderate intellectual disabilities (Consequently, students with regular IQ or severe degree of intellectual disability were excluded). The Spanish standardization of The Kaufman Brief Intelligence Test (K-BIT, Kaufman & Kaufman, 1997) was used to measure the verbal and non-verbal intelligence of the students. This is a brief, individually administered measure of the intelligence of a wide range of individuals spanning the ages of 4–90 years. It is composed of two subtests: Vocabulary (Expressive Vocabulary and Definitions) and Matrices. It takes 15–30 minutes to administer. It was administered by the educational psychologists of the centre *Camí Obert* since it was part of the assessment protocol of the centre. The IQ Composite (a combination of vocabulary and matrices scores) internal consistency coefficient was .98 across ages (.88 to .96). The IQ composite standard

score of the group on average was 67 (SD = 10.08, range=43-88), which corresponds to mild

intellectual disability according to the classification of the DSM-IV-TR Manual (American

Psychiatric Association, 2000) . As scores under 40 represent a severe intellectual disability,

participants scoring below 40 were consequently removed from the sample.

 (b) Reading comprehension level equivalent or superior to the third grade of

elementary school (aged eight to nine). This inclusion criteria was selected to ensure that a

minimum level of reading skill was acquired. This skill was measured through a standardized

test called "Reading Comprehension Test" (ECL, De la Cruz, 1999) for Primary School. The

test is composed of 2 levels, Level 1 for 2nd and 3rd graders and Level 2 for 3rd to 6th graders.

For students with reading difficulties, the manual of the test recommends the use of the Level

1, so this was the version used in our study. The ECL1consistedof three short texts (approx.

average length of 100 words) extracted from real school and literarySpanish books,

accompanied by 17 literal and inferential questions. The application time of the test is 30

minutes. The Cronbach's alpha of the ECL1 is 0.5219. The average direct score obtained was

8.96 (SD = 2.32.), which corresponds to the 50th percentile for the 3rd grade level. This test

was administered by the three first authors of the present study, qualified by a degree in

linguistic (the first one) and psychology (the other two).

Following the performance in these tests, one male participant was withdrawn from

the sample for the score in the ECL1 was too low (direct score = 3 out of 17). The average

age of the final sample was then 19.82 (SD = 2.16, min = 16, max = 24).

There was no main effect of gender for the Anaphor test scores (accuracy), t (26) =

.415; p = .84), so this factor will not be considered in further analyses.

Control group. Twenty-seven (4 males, 23 females) Speech Therapy students aged

between 20 and 25 (M = 21.6, SD = 1.31) from the University of Valencia voluntarily

participated in this experiment as the control group. Reading level and IQ were assumed to

be regular since they were undergraduate students, so baseline measures were not applied in

this group. The 27 undergraduate students received additional course credit for their

participation. In this group, the main effect of gender was neither significant for the Anaphor

test scores (accuracy), t (25) = .834; p = .37, so this factor will not be considered further.

In both groups, students and/or students' parents signed an informed consent before

participating in the study.

Materials. *Anaphor decision task.* For the anaphor decision task, the original stimuli

designed by Gutiérrez-Martínez, García-Madruga, Carriedo, Vila, and Luzón (2005) , also

employed in Elosúa et al.'s (2009) study were used here. These conformed a pool of 84

sentences containing an object pronoun (42 morphosyntactically and 42 semantically related

to the correct answer), followed by two answer choices presented in line, one to the right and

one to the left of the screen. Elosúa et al. et al. followed some guidelines to create the

sentences: (a) each sentence was 10-to-12 words long; (b) the answer choices were two- or

three-syllables high frequency words; (c) only "la" (fem.) and "lo" (masc.) were used as

targets, but not "le" (gender unmarked indirect object) to avoid a possible conflict due to a

well-known misuse of that pronoun in Spanish called "leísmo"; (d) the order of the answer

choices was counterbalanced, so that the correct answer and the distractor occur equally on

the left and the right positions.

The sentence "Carmen **la** (f) encontró en el aula al lado del pupitre" (*Carmen found*

her * *in the classroom, next to the desk)* is given as a real example. In the morphosyntactic

condition, participants had to choose between "carpeta" (f) (folder), and "cuaderno" (m)

(notebook). Both choices match the pronoun semantically, but only the correct answer,

carpeta, also does morphosyntactically. In the semantic condition, the choices would be

"carpeta" (f) (folder), and "profesora" (f) (teacher). This time the concordance is only

established through semantics, as in both choices the gender matches the feminine pronoun but only *folder* makes sense in that context. This test was administered by the three first authors of this research with degrees in linguistic (the first one) and psychology (the other two).

Procedure. Two collective sessions were necessary for the experimental group to perform the tasks: (1) a first session to accomplish the control tests by the experimental group, and (2) a second session, common to both groups, to perform the experimental task itself.

A within-subject design was created to present each participant with a total of 42 non-repeated stimuli, 21 of each kind, in a completely randomized order, by means of a Visual Basic solution created by the experimenters. For methodological reasons we had to change the original between-groups design employed by Elosúa et al., as our sample size was quite smaller than theirs.

In the second session participants were asked to sit on a computer individually and pay attention to the screen. They completed six practice trials followed by the 42 experimental trials. First, an instructions screen appeared to explain the procedure. By pressing the *next* button the first sentence would appear containing an object pronoun in bold. This sentence they should read carefully and, when ready press the *choices* button. The screen then changed to present the two alternatives to be evaluated: two nouns of different gender in the case of morphosyntactically disambiguated pronouns, and two nouns of the same gender for semantic anaphors. The correct answer was the more likely noun to be the antecedent of the pronoun in bold. By clicking in what the participant thought it was the correct choice and then the button *next*, the next stimulus appeared. Re-reading of the sentences was not allowed, but, as explained above, six practice items preceded the actual experimental stimuli in order to guarantee the comprehension of the procedure.

Hypotheses. The main goal of the present study was to obtain comprehensive data about how students with unspecific intellectual disabilities resolve clitic anaphors by measuring reaction times and accuracy in an anaphor decision task.

H1a. Overall accuracy is expected to reach lower levels in both conditions, semantic and morphosyntactic, in the experimental group (readers with unspecific intellectual disabilities) than in the control group. In fact, a ceiling effect is expected for the later according to the results obtained by Elosúa et al., (cf. 2009).

H1b. Accordingly, overall reaction times should be higher in the experimental group than in the control group.

H2. Finally, given some of the antecedents reviewed on low-skilled readers, (Megherbi & Ehrlich, 2005; Yuill & Oakhill, 1988) the experimental group is expected to show a preference for semantics over morphosyntax reflected by increased times and lower accuracy in this latter condition. Following Elosúa et al. (2009), the opposite patter regarding response times is expected for the control group, that is, faster response times in the morphosyntactic than in the semantic condition.

Results

The software employed to administer the Anaphor test also registered participants' choices (accuracy) and times (ms), these defined as the time employed in reading plus answering. Only times of the correct responses are analysed. Mean values and standard deviations for both variables are shown in Table 1.

A first analysis to check the normality and homocedasticity principles revealed a no-normal distribution of the data, so non-parametric analysis were finally performed. For the within-group analysis the Wilcoxon signed-rank test was performed, whereas Mann-Whitney U served for the between-group comparisons.

Additionally, Pearson correlations were performed to explore a possible lineal dependence among the experimental task's results and the control tests, i.e. IQ and reading comprehension. We used an alpha level of .05 for all statistical tests

[Insert Table 1 over here]

Table 1. Experiment 1.Average response times (ms) and accuracy (ms), (*SD*)

Accuracy. First, we tested the hypothesis that accuracy in the Anaphor test in the experimental group was not as good as in the control group (H1). Between-group comparisons yielded some significant results stating that the control group performed significantly better in both the morphological condition ($U = 47, p < .001$) and the semantic condition ($U = 37.5, p < .001$).

Following our H2, significant differences between conditions were expected in the group of students with ID but not in the control group. Within-group analyses revealed a very short difference between the average percentage of correct responses in the morphologically and the semantically disambiguated pronouns, in both the experimental (morphological, $M = 66.96$, *SD*= 16.92; semantic, *M*= 66.21, *SD* = 14.11) and the control group (morphological, $M = 95.11$, *SD* = 5.49; semantic, $M = 95.04$, *SD* = 5.59). This very short difference resulted non-significant (See Figure 1).

[Insert Figure 1 over here]

Figure 1. Experiment 1. Percentage of correct responses in the pronoun resolution task, by group and condition (significant diferences are marked by an asterisk)

Response times. Though both groups followed the same trend, employing longer times in the morphological than in the semantically assigned antecedents, between-group comparisons were yet significant. Thus, in line with H1b, the students with ID employed significantly more time in the morphosyntactic condition than their control peers ($U = 40$, $p <$ .001), as well as in the semantic condition ($U = 28$, $p < .001$).

On the contrary to what happened with the percentage of correct responses, some significant results were found in the within-group analyses for the average time employed in the task. As we had hypothesised (H2), the experimental group dwelled longer when the disambiguation was morphosyntactic ($M = 18370$, $SD = 9033$) than when it was semantic ($M = 15064$, $SD = 6833$), and this difference resulted significant ($Z = -4.21$, $p < .001$). (See Figure 2).However, contrarily to our prediction results from the control group followed the same trend, with longer times in the morphosyntactic ($M = 7235$, $SD = 1592$) than in the semantic ($M = 6788$, $SD = 1560$) condition. Once again, this difference was significant ($Z = -1.99$, p < .05).

An additional analysis was performed to check whether the interference between morphosyntax and semantics was greater in the experimental group than in the control group. To this end, response times in the semantic condition were first deducted from the times in the morphosyntactic condition for each participant. A Mann-Whitney test served for analysing this new variable. As expected, the difference in the experimental group was greater ($M = 3304$, $SD = 3599$) than in the control group ($M = 446$, $SD = 1005$), and this difference resulted significant ($U = 168$, $p < .00$).

[Insert Figure 2 over here]

Figure 2. Experiment 1.Avergage response times (ms), by group and condition (significant

diferences are marked by an asterisk)

Discussion

In this experiment we tested the hypothesis that students with ID present a preference

for following semantic cues when any conflictive information is introduced in a sentence. In

this case, morphosyntax and semantics conflicted in a pronoun resolution task. As expected

(H1a), the chronologically age-matched control group obtained higher scores than the ID

students in both conditions. The times for resolving the anaphor were also hypothesised to be

higher in the experimental group than in the control group (H1b), and this expectation was

also accurate. Finally, a preference of semantics over morphosyntax in response times was

found, but not only in the students with ID (H2), but also in the control group.

A possible explanation for the results of the control group, unexpected given the

abovementioned literature, where an advantage of the morphological condition (with a

pronoun gender cue) over semantic cues (Elosúa et al., 2009) has been found, may be the

design of the task. While in Elosúa et al's experiment the design was between-groups, in our

sit was within-subject, which might have induced a different strategy to resolve the task,

making readers to check both, semantic and morphological cues when both were available

(morphosyntactic condition). This kind of strategy would have supposed a time penalty for

both groups of readers, but especially for the experimental group, with less cognitive

resources to invest in the task.

Experiment 1 did not provide the participants with an ecological environment in which

they could read a complete text in the way they would do it in a normal daily situation.

Presenting the complete text instead of using a sentence-by-sentence reading could also avoid

memory overloads and prevent interferences between the observed variables and, especially, reading times. In 1992 O'Brien and Albrecht designed a methodology aimed to enhance the study of on-line comprehension processes, introducing thus the inconsistency detection paradigm that we use in the present experiment. This paradigm had similarities with previous studies (Baker & Anderson, 1982), but it also supposes substantial changes such as the presentation of the whole text at a time. This study consisted of a set of three experiments, in which Psychology undergraduates read short passages of four to seven sentences. The first sentence introduced the main character and provided information about his or her spatial location (inside/outside). The critical sentence could be either the second or the fifth sentence and was led by a pronoun. The information provided in there could be consistent or inconsistent with the location of the protagonist specified in the first sentence. Apart from recording reading times, a comprehension question was presented after each passage to encourage participants to read for understanding. A *distance* (close vs. distant) variable was introduced in experiment 1, while experiments 2 and 3 did not differ in the stimuli- similar to the close condition of the first experiment-, but in the instructions (adopt the perspective of the protagonist vs. read for comprehension). The results of experiment 1 demonstrated that readers did find the inconsistency and this fact was reflected by an increment of reading times in the inconsistent condition compared with the consistent for both distance conditions. The main effect of consistency was replicated in experiment 3, where specific instructions to adopt the perspective of the protagonist helped participants to detect the ambiguous information and caused them to dwell longer in the inconsistent passages. Oppositely, experiment 2 did not show any significant difference between the two conditions. Though the stimuli were the same as in experiment 3, the instructions were just targeted to reading comprehension what in light of the results, made a great difference.

Since their 1992's study, the design and procedure employed by O'Brien and Albrecht have been extensively used as a mean to obtain on-line and off-line processing measures at a time. This, the inconsistency detection paradigm, gives the researcher the possibility to track reaction (or reading) times, lexical decision times or naming times, among others, and study these data together with off-line comprehension measures such as correct responses to questionnaires. Moreover, researchers can track participants' monitoring skills during reading. It is precisely because of the possibility of obtaining such a complete outlook of comprehension, that the inconsistency detection paradigm has been frequently employed in the study of inference drawing and more relevant to our study, the comprehension of anaphors with ambiguous antecedents. The power of such a paradigm is considerably increased if it is used together with a technology like eye tracking, which on the one hand, renders a large amount of data that allows the researcher to better profit from a single experiment; on the other, eye tracking provides information about assorted online cognitive processes that are otherwise hard to get to. For these two reasons it has been widely exploited (Just & Carpenter, 1980;Daneman & Carpenter, 1983; Traxler & Pickering, 1996; Rayner, 1997; J. Traxler, Bybee, & Pickering, 1997; Pollatsek, Fisher, & Rayner, 1998; Pearlmutter, Garnsey, & Bock, 1999; Traxler, Foss, Seely, Kaup, & Morris, 2000; Van Gompel & Pickering, 2001; Meseguer, Carreiras, & Clifton, 2002; Reichle, Rayner, & Pollatsek, 2003; Hyönä, Lorch, & Rinck, 2003; McDonald & Shillcock, 2003; Hirotani, Frazier, & Rayner, 2006; Pollatsek, Reichle, & Rayner, 2006; Rayner et al., 2006; Dubey, Keller, & Sturt, 2007; Staub & Rayner, 2012; Patterson, 2013) .

A study that profits the inconsistency detection paradigm and eye tracking is that of Rinck et al. (2003) which, by the way, is the one this experiment is inspired in. Their study consisted in two experiments in which they explore the processing of texts with temporal information when this is consistent (or inconsistent) with the situation model built up to then.

Participants were undergraduates who performed a reading task while eye movements (first past fixations, second-pass fixations and regressions) were recorded. The texts to be read consisted of seven sentences; the second sentence introduced the temporal information, while the sixth presented an action either consistent or inconsistent with the information contained in the second sentence. In the first experiment the researchers set the focus on the time course of the processing of consistent vs inconsistent temporal information; in the second, they thoroughly explore how readers behave when those inconsistences are found and, hopefully, they try to recover from them. Additionally, participants in this second experiment had to detect and report the ambiguous information right after reading each text. As expected, readers from the first experiment fixated back the sentence containing the possible disambiguating information more frequently than other areas, yielding more regressions and longer second-pass fixations in the inconsistent condition than in the consistent one. Interestingly, the same results were obtained in the second experiment, but only in those participants who were able to consciously report the inconsistence.

Besides Rinck et al.'s study (2003), another representative example of the use of the inconsistency paradigm together with eye tracking to explore anaphora comprehension is the work of Rayner et al. (2006). In it they presented two experiments, but the one that concerns us here is the second one. This experiment consisted on texts a bit longer (ca. 11 lines) than the ones in the original study by O'Brien and Albrecht, for which consistency (consistent vs. inconsistent) and distance between the antecedent and the anaphor (*close*-ca.12 words- vs. *middle*- ca. 55 words- vs. *far*- ca. 125 words) were manipulated. Instead of analysing the traditional measures of reading times and accuracy, they recorded the eye movements of American adult skilled readers to better understand the moment-by-moment operation of text processing. While no difference between the consistency conditions was encountered in the middle and long distances, the expected behaviour reflecting integration problems (i.e. longer

fixations on the anaphor and more regressions to the antecedent) did appear in the closest

inconsistent condition (ca. 12 words distant). These results suggest that the effect of the

inconsistency diminishes as distance increases, very likely because participants update the

possible antecedents as reading goes on. Rayner et al.'s results are not surprising, since

distance has been previously proven to negatively affect anaphora resolution: as distance

between the anaphor and its antecedent increases, the probability of considering a noun phrase

the correct antecedent decreases (Duffy & Rayner, 1990; Myers & O'Brien, 1998).

Experiment 2 was designed on the basis of the antecedents on the inconsistency detection

paradigm. The rationale, method and results of this second Experiment are described in the

next section.

Experiment 2

The findings from our previous experiment evidenced the problematic of pronoun

resolution experienced by youngsters with ID when there are competing antecedents. Just as

predicted and based on previous literature (Megherbi & Ehrlich, 2005; Oakhill & Yuill,

1986) , there was an increment in readers with ID's response times mainly when the

competing antecedents had to be evaluated and selected following morphosyntactic cues. The

subsequent experiment kept on investigating anaphora resolution, following the work of

Rinck, Gámez, Díaz, and De Vega (2003), this time using short texts instead of the single

isolated sentences of the anaphor test by Elosúa et al. (2009). We put into play a text reading

task from which, on the one hand, data of on-line processing of ambiguous pronoun was

collected by recording eye movements. On the other hand, off-line processing was also tested

through a questionnaire presented after the reading in which readers with ID, a

chronologically age-matched group (young adults without ID), and a reading age-matched

group (typically developing children), had to answer to comprehension questions reflecting

different reading processes and report any inconsistence or ambiguity detected. This
questionnaire consisted in three objective questions targeted to both literal and inferential
information, one more question about the subjective perception of easiness and finally, the
detection of the inconsistency itself. This questionnaire gave us the opportunity to explore the
preference of use of semantic versus morphological cues in each group of readers.

Method

    Participants.

    An experimental group consisting of young students with ID and two control groups
of students without ID participated in this experiment: (1) a chronological age-matched
control group recruited at a mainstream primary school in Valencia (Spain) and (2) a reading
level-matched control group. The experimental group was built up by some of the
participants from the experimental group of the Experiment 1 plus some new candidates also
recruited from the School of vocational Training *Camí Obert (*Valencian Community, Spain).

    Experimental group. The inclusion criteria for participants with ID were similar as
the ones in the previous experiment: (a) borderline to moderate intellectual disabilities and (b)
reading comprehension level equivalent or superior to the third grade of elementary school.
This time, twenty students with borderline to moderate intellectual disability were initially
recruited from *Camí Obert*. Once again, the Kaufman Brief Intelligence Test (K-BIT,
Kaufman & Kaufman, 1997) was employed for measuring IQ, but reading comprehension
level was assessed by means of the text *Penguins (*Cronbach's α= .653), which is one of the
texts contained in the standardized TEC reading test (Vidal-Abarca et al., 2008). The original
probe consists of two texts accompanied by 10 literal and inferential questions each, but we
decided to use only one of the two texts because using the complete version could be

exhausting for the students with ID. Following the K-Bit test battery scores one participant

with a normal IQ was discarded, as well as 2 students with severe intellectual disability. The

low comprehension rates in the reading task caused 4 more students to be pulled out from the

sample (those with scores below the level of the 3rd graders). Additionally, recording errors

were found in one participant's eye tracking data, who had to be withdrawn. Thus, the final

experimental group consisted of 12 students (4 male, 8 female) aged between 16 and 23 years

(M = 19,6, SD = 2,54). The average IQ was 59,25 (SD = 13,3, range = 43-88), which

corresponds with a mild intellectual disability according to the DSM-IV-TR Manual

(American Psychiatric Association, 2000). Regarding the results of the TEC reading test, the

group obtained a mean score of 4,25 (SD = 1,5, range= 3-7), out of 10 questions,

corresponding to a percentile 50 in the 5th graders scale.

   Chronologically age-matched control group. Nineteen students from different

disciplines of the University of Valencia, in Spain voluntarily participated in the study as the

chronological age-matched control group. Due to eye tracking errors three participants had to

be withdrawn, so 16 students remained in the group (5 male, 11 female), aged from 18 to 20

years (M = 19,1, SD = 0,6). For they were all undergraduates, no additional tests were

administered apart from the proper experimental task.

   Reading level-matched control group. The reading age-matched control group first

contained 25 5th graders from a regular Spanish primary school. Three of them were

withdrawn for lacking some of the control tests, 2 because they were not native speakers of

Spanish, which could interfere in the results, one because of the low scoring in the K-Bit, and

1 student more because of low comprehension rates in the TEC (Vidal-Abarca et al., 2008)

reading task (only Text *Penguins* like in the experimental group was applied). This test was

administered by the three first authors of this research qualified with degrees in linguistic (the

first one) and psychology (the other two). Because of recording errors 6 participants more

had to be withdrawn. Thus, a final control group of 12 children (4 male, 8 female) between

11 and 12 years old ($M = 11,6$, $SD = 0,5$) was used for the analyses. The IQ was measured

again with the K-Bit intelligence test battery, obtaining an average of 105.6 ($SD = 15,1$, range

$= 80\text{-}127$). The mean direct score in the TEC reading test for this control group was 5,7 (SD

$= 1,8$, range $= 4\text{-}9$), out of 10 questions, which corresponds to a percentile 60 in the 5$^{\text{th}}$

graders scale. A Student T-test confirmed that the difference between the percentiles obtained

in the TEC in the experimental and the children control group was not significant (t (22) $= -$

2,1, $p = .06$).

The main effect of gender for the accuracy scores of the inconsistency detection task

was not significant for any of the groups: experimental, t (14) $= .636$, $p = .54$; chronological, t

(16) $= .70$, $p = .50$; reading level, t (17) $= .15$, $p = .88$. That means, this factor will not be

taken into account further on.

Participants or their legal tutors signed an informed consent before starting the study.

Materials. *Inconsistency detection task*. Four texts of six sentences were designed

taking Rinck et al.'s (2003) materials as model. Sentences 1, 3, 4 and 6 were neutral, that is

they contained no relevant information for the task though two proper names of different

gender –masculine and feminine- were presented in the first one. Sentences 2 and 5 were the

targets. Sentence 2 stated an action involving the two actants from sentence 1, always being

one of them the subject and the other the indirect object. Sentence 5 was introduced by a

personal pronoun in the syntactic role of the subject, whose antecedent could be either of the

two proper names from sentence 2. (See Appendix A for a text sample).

The study followed a quasi-experimental design with *anaphoric consistence*

(consistent vs. inconsistent) as the only independent variable and *pronoun gender* (masculine

vs. feminine) as a non-experimental variable only used to control for gender bias. Thus, four

different versions were created from each text: consistent (half masculine, half feminine) and inconsistent (half masculine, half feminine). The (in) consistency was formed by matching the pronoun and the information contained in sentence 5 with the subject of the second sentence. In the *consistent* condition the pronoun-subject matching was semantically and morphosyntactically established (e.g. Pedro-he).On the contrary, the *inconsistent* condition happened when that matching was both semantically and morphosyntactically incongruent (e.g. Elena-he). If the pronoun (for instance, "he") did not match the subject of the second sentence ("Elena"), it was possible to attach it to the indirect object ("Pedro"), though the semantics of the context never allowed such a tie.

In addition, text length was controlled between conditions (range: 311-357 words).

The assignment of target texts to experimental conditions was counterbalanced across participants through four lists so that one participant read each experimental story just once and in only one experimental condition. Across participants, each text occurred equally often in each condition. The order of presentation of the four stories was randomized across participants by the SMI Vision design and presentation software *Experiment Center*. After each stimulus, the corresponding set of five comprehension questions appeared on screen one-at-a-time to test for global comprehension and conscious pronominal ambiguity detection.

Comprehension questions. As the goal of the study was to test and examine not only the online but also the offline processing of pronominal anaphora, especially when incongruences in the assignment of an antecedent occur, a set of five questions were designed to be presented after each text: three objective (*non-integrative, antecedent comprehension question,* and *anaphor integration question*) and two subjective questions ("is there anything strange in the text?", "is there anything difficult in the text?") were presented on screen to be answered orally after each text (see Appendix B for a complete questionnaire sample). The

subjective questions were recoded into two new variables: *explicit error detection* and *nature of the information provided*.

   Literal question. The first inquiry referred to the first sentence and thus, it did not imply either the resolution of the anaphora or any interclausal integration (e.g. Where do Pedro and Elena live?).

   Antecedent comprehension question. The second question asked about the information contained in the second sentence (for example, "Who phoned to meet on the beach?"). This time a total comprehension of the action (what happened?) as well as the role of the actants in it (who did what to whom?) was necessary.

   Anaphor integration question. To answer this third question participants should resolve the anaphor, attaching the pronoun in sentence 5 to the most plausible antecedent in sentence 2 (e.g. "who always calls to meet?") From the point of view of the scoring, the consistent condition was clear and simple. The inconsistent condition, on the other hand, was arbitrarily scored on the basis of a morphosyntactic matching in which 1 (one) was assigned to the answer matching the gender of the pronoun and 0 (zero) to the semantic alternative (i.e. Pedro-he = 1; Pedro-she = 0). In other words, using a semantic attachment would sum 0 points, while the morphosyntactic strategy would add 1 point.

   Explicit error detection. Participants answered here the question "do you see something strange/odd in text? This was a simple boolean variable with 0 (zero) as incorrect and 1 (one) as correct response.

   Nature of the information provided. According to the answers to the subjective questions, the kind of information provided by participants was recoded into a three-level categorical variable where 0 (zero) meant no information provided, 1 (one) represented any kind of *morphosyntactic* statement (e.g. "*Pedro* cannot be *her*, that is wrong"), and 2 (two)

was assigned to *semantic* information (e.g. "if Elena knows so much about files, why is Pedro the one resolving the doubt?").

Eye movements. To compute and analyse the dependent variables in eye tracking experiments it is necessary to select the target information in order to group the eye gaze data and thus ease the task of interpreting the results. Each group of eye gaze data is commonly called area of interest, henceforth AoI. In this case, six AoIs were first defined corresponding to the six sentences in the text. After several attempts to obtain any meaningful and explainable results, finally thicker measures had to be employed with no regard to any AoI distinction, mainly because of the irregularities in the eye movements of the participants with intellectual disabilities. Thus, the overall data for each text was finally computed and analysed.

Using eye tracking to study on-line comprehension processes allows the recording of a large amount of data. For this reason, grouping this data into *eye tracking measures* is essential (Rayner, 1997). For the present experiment several measures were computed for both early and late comprehension processes. As early measure, *first pass duration* reflects both the initial word recognition and the lexical processing. It is defined as the sum of the durations of all fixations in a region from first entering the region until leaving it either to the left or the right, given that the region was fixated at least once. As late measures, associated to integrative processes during reading comprehension, the number and duration of the fixations made after leaving the critical area of interest containing the pronoun (sentence 5) were accounted, henceforth *fixations after offset S5* and *duration after offset S5*. This test was administered by the first author of the present research, holding a degree in linguistics and expertise in eye-movements measurement.

Procedure. A total of two sessions were needed to accomplish the tasks. Participants from the experimental group and the reading level-matched control group needed a first collective session in order to complete the K-Bit battery test and the TEC reading test.

In a second individual session participants fulfilled the Inconsistency Detection task. In this second session, the SMI Experiment Center software administered the task while a SMI RED250 stand-alone Eye Tracker was employed to record participants' eye movements throughout the task. Responses to comprehension questions were registered by the experimenter.

After giving the participant instructions to seat comfortably and to stay as quiet as possible during the task, the eye tracker was adjusted for optimal tracking. Participants were calibrated by a standard 9-point grid followed by another verification 9-point grid. They were instructed to read a set of short texts carefully and answer the questions verbally. The task was self-paced, so they could decide when to skip the screen and move on to the next question or text. Eye movements were recorded only during the reading of the text, not the questions.

Before each story, a fixation cross appeared centered on the screen surrounded by an invisible trigger-zone, so that after a 1000 ms gaze inside the boundaries the software automatically triggered the next stimulus. The next screen after reading the text appeared by pressing any key. At this point the student should read the question on screen and answer verbally. A pause could be done whenever the participant asked for it. They completed 2 practice trials followed by the experimental items. The experimental trials did not start until the experimenter verified that the procedure had been understood.

Hypotheses. H1. Following the antecedents reviewed (Fajardo et al., 2013; 2014), students with ID are expected to experience problems in the inconsistent condition with the two inferential but not with the literal question. The ability to detect the inconsistency should be low, as it depends to a great extent on their metacognitive skills and these have been proven to be irregular in individuals with ID (Doğanay & Özmen, 2014). For the same reason, no difference between conditions is expected for the times and number of fixations after encountering the conflictive pronoun.

H2. The chronologically age-matched group would also perform equally in both conditions of the literal question since it is not affect by the anaphor inconsistency, but not in the inferential ones where lower accuracy scores are expected in the inconsistent than in the consistent condition. Overall scores are expected to be higher than in the experimental or in the children groups. For metacognitive skills are supposed to be normally developed, detection will also be better in this group than in their peers', and they will also dwell longer and with more fixations after reading the conflictive information than the other groups.

H3. At last, the reading age-matched group should obtain similar results to the ID students: lower accuracy in the inferential questions than in the literal question, in overall under the chronologically age-matched group scores. The inconsistency will not affect them greatly because they are supposed to be still developing their metacognitive and literacy skills, so times and number of fixations after the reading of the incongruent information should be similar in both experimental conditions.

Results

For the analysis of questions no participant was excluded because of recording errors, as these only appeared in eye tracking. Instead, the reading skills and IQ did serve as filters. The sample here was greater then, constituted by 16 participants in the experimental group, 18

in the chronologically age-matched readers' and 19 in the reading level-matched group. Non-parametric analyses, Mann-Whitney U for between-group comparisons and Wilcoxon matched pairs for within-group analyses, were performed to analyse the objective questions and eye tracking data. The variable *explicit error detection* was analysed by means of a Student-T, and finally, a Fisher's exact test served to explore the variable *nature of the information provided*. Data from the questionnaire is shown in Figure 3 and Figure 4.

Regarding eye movements, a filtering of outliers was performed prior to the analyses, considering outliers the times two standard deviations over or under the mean of each participant and condition. These values were replaced by the average time of the participant per fixation and they constituted a 3.5% of the values.

An alpha level of .05 was employed for all statistical tests.

Experimental group. No difference between conditions was encountered in the *literal question* or the *antecedent comprehension question*. Regarding the *anaphor integration question*, the ID group obtained higher scores in the consistent ($M = 84$) than in the inconsistent condition ($M = 34$), $Z = -3.56$, $p < .001$. The *nature of the information provided* did not yield any significant difference, though "nothing" was reported more times than "morphosyntactic" or "semantic" (nothing, $N = 21$ vs. morphosyntactic, $N = 5$, semantic $N = 6$). Regarding eye movements, nor *first pass times*, *number of fixations after the offset of sentence 5*, or *duration of the fixations after the offset of sentence 5* reported any difference between the consistent and the inconsistent condition.

Chronologically age-matched control group. Again, results from the *literal question* reported no significance between conditions, and these were also close to a ceiling effect. This time, results from the *antecedent comprehension question* did not follow the same trend as in the experimental group, with significantly higher scores in the consistent ($M = 97$) than in the inconsistent condition ($M = 50$), $Z = -3.69$, $p < .001$. The *anaphor integration question*,

on the other hand, did so, showing higher scores in the consistent condition ($M = 100$) than in the inconsistent ($M = 47$), and resulting this difference significant, $Z = -3.95$, $p < .001$. Fischer's exact test revealed a significant effect of *kind of information* in the *nature of the information provided* variable, $p < .001$. Besides, the trend was different as the other groups' (nothing, $N = 8$, morphosyntactic, $N = 25$, semantic $N = 3$). In line with the experimental group, eye tracking measures did not yield any significant result.

Reading age-matched control group. Results from the *literal question* were unexpected here, for children performed significantly better in the consistent ($M =100\%$) than in the inconsistent condition ($M = 87\%$), $Z = -2.24$, $p < .03$. Following the same trend as the chronologically age-matched group, the *antecedent comprehension question* in its consistent condition yielded a higher percentage of correct answers ($M = 95$) than the inconsistent condition ($M = 76$), and this difference was significant, $Z = -2.65$, $p < .01$. The same happened in the *anaphor integration question*, with a greater preference for semantic cues in the consistent condition ($M = 95$) compared to the inconsistent ($M = 53$), and resulting the difference significant, $Z = -2.24$, $p < .001$. When it comes to *nature of the information provided*, results in the children's group were similar to the students with ID (nothing, $N = 21$, morphosyntactic, $N = 10$, semantic $N = 7$), showing no significant effect of *kind of information* at all. Once again, eye tracking data reported no difference between conditions.

[Insert Figure 3 over here]

Figure 3. Experiment 2. Average scores in the questionnaire, inconsistent condition by group (*average morphosyntactic matching)

Between-group. As we had hypothesised, no difference among groups was encountered in the *literal question* (see Figure 3 for a comparative view of the scores in all

questions for the inconsistent condition). However, some differences appeared in the

performance for the *antecedent comprehension question*. Children performed better than the

chronologically age-matched group in this question, $U = 91$, $p < .02$, but these results did not

replicate when comparing to the ID students. The *anaphor integration question*, by its part,

did not report any significant difference among groups, as all of them performed similarly:

over 80% semantics in the consistent condition (morphosyntactic cues followed 20% of the

times), and 60% in the inconsistent (morphosyntactic preference 40% of the times).

Regarding the *explicit error detection*, just as expected, the chronologically age-matched

group performed significantly better than the experimental group, t (32) = -3.30, $p < .003$, and

the children, t (35) = -3.24, $p < .004$, (see Figure 4). An interaction between *group* and *kind

of information* was found, $p < .001$, very likely because of the different results of the

chronologically age-matched group compared with the experimental group and the reading

level-matched group, concretely in what to the morphosyntactic category refers.

[Insert Figure 4 over here]

Figure 4. *Experiment 2. Kind of information reported (N) in the inconsistent condition by

group*

The *duration of the first pass* was similar across groups, so analyses did not return any

significance. The *number of fixations after the offset of sentence 5*, on the other hand, yielded

some significant results. As we had hypothesised, normative adults made more fixations after

the pronoun than the experimental group in both the consistent (chronologically age-matched,

$M = 33.97$; experimental group, $M = 11.33$) and the inconsistent conditions (chronologically

age-matched, $M = 33.75$; experimental group, $M = 21.21$). In both cases the difference was

significant (consistent, $U = 23.5$, $p = .001$; inconsistent, $U = 52$, $p = .04$). Additionally, the

children were encountered to refixate the text more than the ID students in the consistent condition (reading age-matched, $M = 32.58$; experimental group, $M = 11.33$), and this difference was significant, $U = 26$, $p < .01$. Regarding the *duration of the fixations after the offset of sentence 5*, some unexpected results were obtained, as the chronologically age-matched group dwelled significantly longer in re-reading the text when no inconsistence was introduced than the ID students (chronologically age-matched, $M = 6095$ ms; experimental group, $M = 2545$ ms), $U = 33.5$, $p < .01$. The same happened in the reading age-matched group in comparison with the experimental group (reading age-matched, $M = 6882$ ms; experimental group, $M = 2545$ ms), $U = 32$, $p < .03$.

Discussion

In this second experiment we re-tested the hypothesis of the semantic vs. morphosyntax advantage in pronoun resolution by readers with ID, this time presenting a pronominal anaphor in a text in an inconsistency detection task.

For the experimental group we expected low accuracy in the inconsistent condition of the inferential questions (H1) but this was only partially accomplished. These results were obtained only in the *anaphor integration question*, showing a preference for the semantic pronoun-antecedent matching, but they performed equally well in both experimental conditions of the *antecedent comprehension question*. A possible explanation is a failure in their metacognitive skills, so that they do not notice the ambiguity answering this question in both conditions. Similarly, both eye movements and *explicit error detection* scores showed this same poor metacognitive performance. Regarding the kind of *information provided* about the nature of the ambiguity, the students with ID reported "nothing" in most of the cases, while "morphosyntactic" and "semantic" obtained rather similar results, showing thus no preference for one or another strategy.

The chronologically age-matched control group was expected to show clear differences between the consistent and the inconsistent conditions in the two inferential questions (H2), and so did they, with a better performance in both cases in the consistent condition. In other words, they were actually sensitive to the anaphoric inconsistency. Regarding the prediction about the preference of a semantic strategy, scores in the *anaphor integration question* were close to 50%, reflecting no preference of such a strategy over the morphosyntactic one. *Explicit error detection* was high (ca. 70% correct detections), but eye movements did not show any change in the reading strategy from the consistent to the inconsistent condition. About the *nature of the information provided*, the control adults reported the inconsistency far more frequently to be a "morphosyntactic" issue than "semantic" or "nothing". This result is consistent with the other variables, supporting the hypothesis that normative adults possess a better metacognitive ability.

Following our hypothesis H3, the children group should have had similar results to the ID students, reflecting low metacognitive skills in any case. Surprisingly, they obtained higher scores in the consistent condition not only in the *antecedent comprehension question*, but also in the *literal question*. As children also obtained scores close to 50% in the *anaphor integration question*, we draw the same conclusion here as for the chronologically age-matched group: they show no preference for any of the two tested strategies. As expected, *explicit error detection* was very low and eye movements were similar in both experimental conditions.

Comparing the groups, the chronologically age-matched control group also demonstrated to owe better metacognitive skills. Even though no significant result was obtained in the between-group comparisons for the *literal question* or the *anaphor integration question*, they obtained lower scores in the *antecedent comprehension question*, very likely due to the impact of the ambiguity. Children seem to have realised about the inconsistency, but they probably

evaluate it as not so important, so they could easily answer the question, which asked about information contained in sentence 2 (no integration of the anaphor was needed). The adults control group, on the contrary, did completely notice the ambiguity, so they made it extensive to the comprehension of the antecedent, experiencing problems to answer this question. Conscious detection was also better in this group. When it comes to eye movements, no group spent more time than the others in any condition. However, both children and normative adults made more refixations after reading the last sentence than the ID students, meaning that the default reading strategy of both control groups was qualitatively different than the ID students'.

General discussion

For this study we designed two experiments to explore anaphora resolution under conflictive circumstances. In Experiment 1 we adapted the anaphora test of Gutiérrez-Martínez et al. (2005) and Elosúa et al. (2009) to test how students with ID and a chronologically age-matched group of undergraduates performed a task involving anaphors, morphosyntax and semantics. Participants should decide, between two answer choices, which word a clitic pronoun contained in a given sentence referred to. Half of the times, the correct answer could be unravelled through a gender cue, while the other half it was the semantics of the context and the word itself what served to that end. Both groups performed equally well in the two experimental conditions. However, the times employed in resolving the morphosyntactic anaphor were significantly higher than the semantic condition, in both groups. As expected, the normative adults performed in overall better than the ID students.

Experiment 1 inspired us to continue our investigation in the line of the contrast morphosyntax vs. semantics. Hence, we designed Experiment 2 inspired in Rinck et al. (2003) and the inconsistency detection paradigm to assess both on-line and off-line processing

of anaphoric pronouns and detection of binding inconsistences. Participants read four texts whose second sentence contained two proper names of the same gender (in subject and indirect object roles) while the fifth sentence was introduced by a pronoun bound to one of those names, half the time the subject, half the time the object (in this case, overall semantics resulted ambiguous). To avoid gender bias, as it has been proven to influence pronoun resolution (Arnold et al., 2000), half the names were masculine and half feminine. By recording eye movements we wanted to obtain on-line evidence about the recognition and integration processes of ambiguous and unambiguous pronouns. An additional questionnaire would allow us to collect data about participants' awareness of this ambiguity, their objective disambiguating accuracy and their literal text comprehension.

The results obtained in Experiment 1 were similar to Elosúa et al.'s in accuracy, as no difference between the experimental conditions was found in both groups. Though overall mean accuracy was lower in the ID students than in the control group, no difference at all was found. These results are consistent with the fact that the stimuli were originally conceived to serve in an operative memory test, so they were created intentionally easy for young adults.

Finding an explanation to processing times seems trickier. While Elosúa et al. obtained evidence supporting the superiority of morphosyntactic analysis, reflected by shorter processing times, our results go in the opposite direction. The reason why our students rely on semantics instead of morphosyntax may lay on the design itself. Elosúa et al. manipulated the experimental conditions between-groups, meaning that every participant was exposed to only one condition, either semantic or morphosyntactic. Our design, on the other hand, was a within-group, so that every participant read the same number of items of each condition and the items of each condition were mixed in such a way that within the same trial block participants resolved both morphosyntactic and semantic trials presented in a random order. This fact could be having an impact on the results in what to strategies refers: as they realize

that gender cues are only sometimes present, they decide to employ a semantic strategy by default. When the anaphor is to be semantically disambiguated, response times are shorter than Elosúa et al.'s control group. On the contrary, when they are supposed to employ morphosyntactic cues but semantics are used instead, a penalty in time appears as switching from semantics to morphosyntax supposes a delay.

As the participants in Elosúa et al.'s study were young adults with no explicit literacy problem, it is understandable that the results obtained by our ID students remain unexplained from the same perspective. However, the results obtained in our Experiment 1 for this group are similar to Megherbi and Ehrlich's (2005). They observed that only skilled readers were able to take advantage of morphosyntax when gender cues are available besides semantics. Low-skilled children, on the other hand, preferred to resolve the anaphor following the semantics of the verb instead of the morphosyntactic cues. These results could explain why the children and ID students in our experiment seem to prefer the semantic disambiguation as default strategy, even though taking advantage of a surface element such as gender signs could aid a faster resolution of the anaphor (Cacciari et al., 1997).

For Experiment 2 we argued that students with ID would experience problems with the two inferential questions (H1), but the results did not completely accomplish this expectation. In fact, only the question asking directly about the integration of the anaphor resulted more intricate in its inconsistent condition. Consequently with our predictions, though the inconsistency did cause some problems in the resolution of the anaphor (34% correct responses in the *anaphor integration question*), students with ID neither notice it consciously (22% correct detections) nor changed their reading strategy to re-read the incongruent parts of the text (no difference between conditions in eye movements). These results were in overall what we expected, as in the literature review we had found enough antecedents reporting the limited metacognitive skills of individuals with ID (e.g. Doğanay &

Özmen, 2014). Besides, given the low reading comprehension level of the group, it was anyway reasonable to anticipate a behaviour proper from low-skilled readers in terms of a fairly detection and a poor reading strategy.

The chronologically age-matched control group encountered some problems in the two integrative questions, just as we hypothesised (H2). Given the nature of the *antecedent comprehension question* and the scoring criteria of the *anaphor integration question,* we can assert that these scores, in both cases close to 50%, reflect no preference of semantics or morphosyntax, for they seem to use both equally. Once again, this finding is consistent with Megherbi and Ehrlich's, who found skilled readers to take advantage of both cues in a pronoun resolution task. Precisely for this reason, it is a bit puzzling that participants from this group reported the inconsistency as a morphosyntactic far more frequently than a semantic issue. It may be that they consciously know the grammar rules of the language and after detecting that the problematic of the text lays on the pronoun (67% correct detections), they get to the conclusion that grammar should be responsible of such incongruence. A bit surprising is the lack of effect of the inconsistency in eye movements, because this group was expected to be affected by it but, on the contrary, they followed the same reading strategy in both kinds of texts.

Regarding the reading age-matched results, our hypothesis 3 was again quite accurate. Children showed some problems with the inferential questions when the anaphor was not consistent within the context. However, they also performed much better in this condition when a simple literal question was made. For any reason, they seem to encounter the text globally inconsistent when an inconsistence is introduced, even though this was supposed to affect only to the integration of the sentences 2 and 5 (*anaphor integration question*). Detection skills were very low (24% correct detections), similarly to the ID students', just as we had expected. Again, the low detection rates together with the absence of effect of the

inconsistency in the eye tracking data reflect impoverished metacognitive skills that prevent children from performing an adequate reading strategy.

When we compared the scores and behaviour of the three groups some outstanding results came to light. Very interestingly, a main effect of inconsistency was encountered in the *antecedent comprehension question* for both control groups. Even though they performed significantly better in the consistent condition, children's accuracy was higher than normative adults'. This apparently surprising result might not be so rare, given that the questionnaire was presented once the text had already been completely read and the chronologically age-matched group, who in the light of the results have greater inconsistency detection skills, might be more affected by the ambiguity than the children.

When it comes to the explicit error detection, the chronologically age-matched group did perform significantly better than the children and the ID students, what completely matches our expectations.

Taking all the results together, we could argue that children as well as normative adults are more likely to re-read the text than the students with ID as a default strategy, maybe to check whether the situation model they have constructed through the first reading is trustworthy (Baker & Anderson, 1982) . Nonetheless, when an inconsistency is introduced only the chronologically age-matched seem to detect it, making more refixations than the experimental group. This behaviour posits the use of different reading strategies in the typically developing individuals and youngsters with intellectual disabilities, although apparently the normative adults' is the most efficient strategy, for they are able to detect the conflictive information more frequently than the other participants.

Regarding the *nature of the information provided* and taking the results of Experiment 1 into account, a higher count of semantic explanations were to be expected in the ID students and possibly the children, but not in the chronologically age-matched control group, where no

preference for a certain strategy was hypothesised. This expectation was not completely achieved. Instead, a substantial count of *morphosyntactic* explanations were given by the chronologically age-matched group in comparison with the *semantic* or the *nothing* categories, meaning that they regard the inconsistency as a question of morphosyntax and not so much of semantics. In the children and the ID students groups the trend was quite different: very few times they pointed to semantics or morphosyntax, but they reported much frequently *nothing*. We argue that the contradiction between these results and the questions' might be explained again by means of the impoverished metacognitive skills.

At last, we also hypothesised better detection results in the inconsistent condition in terms of a higher count of refixations and longer refixation times in the normative young adults in comparison with the children and ID students. What we observed was that, in fact, the adults without ID made more fixations in second and subsequent readings than their peers with ID. Also children did, but only in the consistent condition. On the one hand, this result is intended in the same line as the other variables, as a reflection of better metacognitive skills in non-ID adults than in ID students. On the other, it states that the strategies used by the three groups tested are qualitatively different: while children and adults tend to re-read the text, either including any ambiguity or not, the ID students do not check their own model of the situation. Taking this result isolated, it does not necessarily mean that they do not notice the inconsistency, as previous research has already given account of the "laziness" of low-skilled readers, who do not modify the gist of their models even though they detect some inconsistent information (van Oostendorp, Otero, & Campanario, 2002; Otero & Campanario, 1990; Otero & Kintsch, 1992) . However, together with the questionnaire results, we should claim that the strategy of our ID students is poorly efficient, because they rarely review their model of the situation, and they hardly ever manage to detect the inconsistency.

Conclusions

The results obtained in these experiments give us a new insight into how the detection of inconsistencies works in the observed populations. We have seen that there exist little and isolated differences between Spanish 5th graders and normative young adults in what to reading strategies refers. Children are, however, not completely competent in detecting the inconsistency, as those strategies only resulted successful in the case of the young adults. Factually, youngsters with intellectual disabilities and children obtained similar off-line results despite the differences in the on-line processing. This is an interesting result from the point of view of a reading intervention: despite the isolated differences in the processing of ambiguous pronouns that can be found throughout this research, it seems that the level of awareness is rather similar in children and young adults with intellectual disabilities, so a proper intervention to train ID students in reading strategies might be worthy. In the framework of text design, the conclusions from this research come to say that including pronominal anaphora is not strictly a *bad practice* when text comprehension is to be enhanced, for the participants in our studies seem to have problems only when the antecedents of these anaphors are not clear or they are presented in an unnatural reading context (like the isolated sentences of Experiment 1) where there is few semantic information available to build the text representation (O'Brien & Albrecht, 1992) .

Despite reading skills are central in curricula of special education or inclusion situations, literacy acquisition is still limited for students with ID as recent studies have shown (Ratz, & Lenhard, 2013; van Wingerden, Segers, van Balkom, & Verhoeven, 2014). Thanks to the information provided by the present study we have learnt that training on the use of metacognitive skills should be incorporate to these curricula, instead of training the mere isolated decoding skills during reading comprehension. Actually, teaching to use metacognitive strategies like the activation of prior knowledge have been successful to

increase inference making during reading in students with ID (Morgan, Moni, & Jobling, 2004). However, it also seems that the processing of morphological cues does not occur automatically in people with ID, who prefer to use semantic cues even though this could eventually be reducing the available memory resources needed to put other metacognitive skills into play. This reduction of the available memory resources during reading could be especially relevant for students with ID due to their low short term memory capacity (Van der Molen, Henry & Van Luit, 2014). Therefore, an alternative or even parallel strategy to increase reading comprehension is suggested to improve morphosyntactical skills. This alternative to traditional training programmes on reading could have a direct beneficial effect on reading comprehension but could also liberate resources devoted to metacognition. This solution could be especially interesting for some aetiologies of ID like Down syndrome, which seems to present a specific delay on syntax acquisition (Fowler, 1990).

References

American Psychiatric Association. (2000). *DSM-IV. Diagnostic and Statistical Manual of Mental Disorders 4th edition TR.* American Psychiatric Association.

Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*, *22* (4), 527–565.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition*, *76* (1), B13–B26.

Baker, L., & Anderson, R. I. (1982). Effects of Inconsistent Information on Text Processing: Evidence for Comprehension Monitoring. *Reading Research Quarterly*, *17* (2), 281.

Cacciari, C., Carreiras, M., & Barbolini, C. (1997). When Words Have Two Genders: Anaphor Resolution for Italian Functionally Ambiguous Words. *Journal of Memory and Language*, *37* (4), 517–532.

Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal*, *11*, 489–503.

Cain, K., Oakhill, J., Barnes, M. A., & Bryant, P. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition*, *29* (6), 850–9.

Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The Use of Heuristic Strategies in the Interpretation of Pronouns. *Journal of Psycholinguistic Research*, *19* (4), 246–264.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9* (4), 561–584.

De la Cruz López, M. V. (1999). Evaluación de la comprensión lectora: ECL. Madrid: Tea.

Department of Education (2008). PCPI Regulating Ordinance (Publication No. 5790/23.06.2008). Valencian Community, Spain.

Doğanay Bilgi, A., & Özmen, E. R. (2014). The Impact of Modified Multi-component Cognitive Strategy Instruction in the Acquisition of Metacognitive Strategy Knowledge in the Text Comprehension Process of Students with Mental Retardation. *Educational Sciences: Theory & Practice*, *14* (2), 707–715.

Dubey, A., Keller, F., & Sturt, P. (2007). The Effect of Phonological Parallelism in Coordination Evidence from Eye-tracking. In *Proceedings of the 2nd European Cognitive Science Conference* (pp. 131–136).

Duffy, S. A., & Rayner, K. (1990). Eye Movements and Anaphor Resolution: Effects of Antecedent Typicality and Distance. *Language and Speech*, *33* (2), 103–119.

Ehrlich, M.-F., Remond, M., & Tardieu, H. (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in metacognitive monitoring. *Reading and Writing: An Interdisciplinary Journal*, *11*, 29–63.

Elosúa, M. R., Carriedo, N., & García-Madruga, J. A. (2009). Dos nuevas pruebas de Memoria Operativa de Anáforas. *Infancia Y Aprendizaje*, *32* (1), 97–118.

Fajardo, I., Ávila, V., Ferrer, A., Tavares, G., Gómez, M., & Hernández, A. (2014). Easy-to-read texts for students with intellectual disability: linguistic factors affecting comprehension. *Journal of Applied Research in Intellectual Disabilities : JARID*, *27* (3), 212–25.

Fajardo, I., Tavares, G., Ávila, V., & Ferrer, A. (2013). Towards text simplification for poor readers with intellectual disability: when do connectives enhance text cohesion? *Research in Developmental Disabilities*, *34* (4), 1267–79.

Fowler, A. E. (1990). Language abilities in children with Down syndrome: Evidence for a specific syntactic delay. *Children with Down syndrome: A developmental perspective*, 9, 302-328.

Frederiksen, J. R. (1981). Understanding anaphora: Rules used by readers in assigning pronominal referents. *Discourse Processes*, *4* (4), 323–347.

Garvey, C., Caramazza, A., & Yates, J. (1975). Factors influencing assignment of pronoun antecedents. *Cognition*, *3* (3), 227–243.

Gelormini-Lezama, C., & Almor, A. (2011). Repeated Names, Overt Pronouns, and Null Pronouns in Spanish. *Language and Cognitive Processes*, *26* (3), 437–454.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, Names, and the Centering of Attention in Discourse. *Cognitive Science*, *17* (3), 311–347.

Gutiérrez-Martínez, F., García-Madruga, J.-A., Carriedo, N., Vila, J.-O., & Luzón, J.-M. (2005). Dos pruebas de amplitud de memoria operativa para el razonamiento (Two working memory measures for reasoning). *Cognitiva*, *17* (2), 183–207.

Hirotani, M., Frazier, L., & Rayner, K. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, *54* (3), 425–443.

Hyönä, J., Lorch Jr, R. F., & Rinck, M. (2003). Eye Movement Measures to Study Global Text Processing. In *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (pp. 313–334). Amsterdam.

Jakubowicz, C. (1991). L'acquisition des anaphores et des pronoms lexicaux en français. In J. Guéron & J. Y. Pollock (Eds.), *Grammaire générative et syntaxe comparée* (pp. 229–253). Paris: Editions du CNRS, Sciences du Langage.

Järvikivi, J., van Gompel, R. P. G., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: contrasting the first-mention and subject-preference accounts. *Psychological Science*, *16* (4), 260–4.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87* (4), 329–354.

Kaufman, A. S., & Kaufman, N. L. (1997). *Test breve de inteligencia de Kaufman (K-BIT).* Madrid: TEA.

Long, D. L., & De Ley, L. (2000). Implicit Causality and Discourse Focus: The Interaction of Text and Reader Characteristics in Pronoun Resolution. *Journal of Memory and Language*, *42* (4), 545–570.

McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, *14* (6), 648–52.

Megherbi, H., & Ehrlich, M.-F. (2005). Language Impairment in Less Skilled Comprehenders: The On-Line Processing of Anaphoric Pronouns in a Listening Situation. *Reading and Writing*, *18* (7-9), 715–753.

Meseguer, E., Carreiras, M., & Clifton, C. (2002). Overt reanalysis strategies and eye movements during the reading of mild garden path sentences. *Memory & Cognition*, *30* (4), 551–561.

Morgan, M., Moni, K., & Jobling, A. (2004). What's it all about? Investigating reading comprehension strategies in young adults with Down syndrome. *Down Syndrome Research and Practice*, 9 (2), 37-44.

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *26* (2-3), 131–157.

O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a

    mental model. *Journal of Experimental Psychology. Learning, Memory, and*

    Cognition, *18* (4), 777–84.

Oakhill, J., & Yuill, N. (1986). Pronoun Resolution in Skilled and Less-Skilled

    Comprehenders: Effects of Memory Load and Inferential Complexity. *Language and*

    Speech, *29* (1), 25–37.

Oostendorp, H. van, Otero, J., & Campanario, J. M. (2002). Conditions of updating during

    reading. In M. M. Louwerse & W. van Peer (Eds.), *Thematics: Interdisciplinary*

    Studies (pp. 55–76). Amsterdam: John Benjamins Publishing Co.

Otero, J. C., & Campanario, J. M. (1990). Comprehension evaluation and regulation in

    learning from science texts. *Journal of Research in Science Teaching*, *27* (5), 447–

    460.

Otero, J., & Kintsch, W. (1992). Failures to Detect Contradictions in a Text: What Readers

    Believe Versus What They Read. *Psychological Science*, *3* (4), 229–235.

Patterson, C. (2013). Discourse Coherence in Pronoun Resolution. *Discours*, *12* (12).

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement Processes in Sentence

    Comprehension. *Journal of Memory and Language*, *41* (3), 427–456.

Perovic, A. (2006). Syntactic deficit in Down syndrome: More evidence for the modular

    organisation of language. *Lingua*, *116* (10), 1616–1630.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: exploring

    the interface between cognition and eye-movement control. *Cognitive Psychology*, *52*

    (1), 1–56.

Rademaker, A., & Haeusler, E. H. (2008). Toward Short and Structural ALC-Reasoning

    Explanations: A Sequent Calculus Approach. Proceedings of Brazilian Symposium on

Artificial Inteligence. Advances in Artificial Intelligence - SBIA 2008. 167-176.

Springer: Berlin.

Ratz, C., & Lenhard, W. (2013). Reading skills among students with intellectual disabilities. *Research in developmental disabilities*, 34 (5), 1740-1748.

Rayner, K. (1997). Understanding Eye Movements in Reading. *Scientific Studies of Reading*, *1* (4), 317–339.

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye Movements as Reflections of Comprehension Processes in Reading. *Scientific Studies of Reading*, *10* (3), 241–255.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105* (1), 125–57.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: comparisons to other models. *The Behavioral and Brain Sciences*, *26* (4), 445–76; discussion 477–526.

Rinck, M., Gámez, E., Díaz, J. M., & De Vega, M. (2003). Processing of temporal information: evidence from eye movements. *Memory & Cognition*, *31* (1), 77–86.

Santiuste, V. (1997). Adquisición de anáforas reflexivas y de pronombres desde la perspectiva de la teoría de principios y parámetros. *Revista de Filología Románica*, *1* (14), 445–454.

Staub, A., & Rayner, K. (2012). *Eye movements and on-line comprehension processes*. *Language and Cognitive Processes* (Vol. 4, pp. SI21–SI49). Oxford University Press.

Traxler, M. J., Bybee, M. D., & Pickering, M. J. (1997). Influence of Connectives on Language Comprehension: Eye tracking Evidence for Incremental Interpretation. *The Quarterly Journal of Experimental Psychology Section A*, *50* (3), 481–497.

Traxler, M. J., Foss, D. J., Seely, R. E., Kaup, B., & Morris, R. K. (2000). Priming in sentence

　　processing: intralexical spreading activation, schemas, and situation models. *Journal*

　　of Psycholinguistic Research, *29* (6), 581–95.

Traxler, M. J., & P, M. J. (1996). Case-marking in the Parsing of Complement Sentences

　　Evidence from Eye Movements. *The Quarterly Journal of Experimental Psychology*,

　　49A (4), 991–1004.

Van der Molen, M. J., Henry, L. A., & Van Luit, J. E. H. (2014). Working memory

　　development in children with mild to borderline intellectual disabilities. *Journal of*

　　Intellectual Disability Research, 58 (7). 637-650.

Van Gompel, R. P. G., & Pickering, M. J. (2001). Lexical guidance in sentence processing: A

　　note on Adams, Clifton, and Mitchell (1998). *Psychonomic Bulletin & Review*, *8* (4),

　　851–857.

Van Wingerden, E., Segers, E., van Balkom, H., & Verhoeven, L. (2014). Cognitive and

　　linguistic predictors of reading comprehension in children with intellectual disabilities.

　　Research in developmental disabilities, 35 (11), 3139-3147.

Vidal-Abarca, E., Gilabert, R., Martínez, T., Sellés, P., Abad, N., & Ferrer, C. (2007). Test de

　　Estrategias de Comprensión (TEC). Madrid: Instituto Calasanz de Ciencias de la

　　Educación.

Yuill, N., & Oakhill, J. (1988). Understanding of anaphoric relations in skilled and less skilled

　　comprehenders. *British Journal of Psychology*, *79* (2), 173–186.

APPENDIX 1

English translation of a sample text used in the experiment

1.	Pedro (m) and Elena (f) are two friends who live in Alicante
2a	<u>Consistent</u>:
.	*Elena phoned to Pedro to spend the day on the beach
2b	<u>Inconsistent</u>:
.	*Pedro phoned to Elena to spend the day on the beach
3.	Everything was almost ready, but the beach umbrella was really dirty and broken
4.	In a shop on the beach Elena and Pedro bought a new beach umbrella
5.	<u>Target sentence</u>:
	She always calls her friends to meet
6.	In summer they will go to the beach more frequently

Note—Each participant read either Sentence 2a or 2b. (m) : masculine; (f) : feminine.

APPENDIX 2

English translation of the sample set of questions for the text in Appendix 1

Literal question	Antecedent comprehension question	Anaphor integration question	Subjective Question 1	Subjective Question 2
Where do Pedro and Elena live?	Who phoned to meet on the beach?	Who always calls to meet?	Is there anything strange in the text? (What?)	Is there anything difficult in the text? (What?)

ANSWER CHOICES

In Alicante	Elena	Elena		
On the beach	Pedro	Pedro	(OPEN)	(OPEN)
In Valencia	Nobody	Nobody		